Technical Report - Ulmer Informatik-Berichte No. 99-01
Department of Computer Science, University of Ulm, Germany.

A Comparison of Multimedia Document Models Concerning Advanced Requirements

Susanne Boll, Wolfgang Klas, Utz Westermann
Databases and Information Systems (DBIS),
Computer Science Department, University of Ulm, Germany
{boll, klas, westermann}@informatik.uni-ulm.de

Abstract

Existing multimedia document models like HTML, MHEG, SMIL, and HyTime lack appropriate modeling primitives that meet specific requirements given by advanced multimedia information system applications. In traditional multimedia applications, multimedia document models just had to cope with the modeling of the temporal, spatial, and interactive course of a multimedia presentation. However, we seriously question whether existing models fit the needs of next generation multimedia applications that bring up requirements like reusability of multimedia content in different presentations and contexts, and adaptation to user preferences. In this paper, we motivate and present new requirements stemming from advanced multimedia applications and the resulting consequences for multimedia document models. Along these requirements, we discuss HTML, HyTime, MHEG, SMIL, and ZyX, a new model that has been developed with special focus on reusability and adaptation. The analysis and comparison of the models show the limitations of existing models, point the way to the need for new flexible multimedia document models, and throw light on the many implications on authoring systems, multimedia content management, and presentation.

Keywords: Multimedia document model, multimedia databases, educational medical applications.

1 Introduction

The initial requirements to multimedia documents were the modeling of the temporal and spatial course of a multimedia presentation. Soon the importance of interactivity for multimedia applications was understood and interaction modeling formed an additional requirement. To offer suitable support for multimedia applications the development of multimedia document models began. On the one side standardization activities started and on the other side commercial tools were evolving. The development and the passing of standards took quite a long time, while in the meantime very sophisticated commercial multimedia authoring tools came to the market that support their own *proprietary* format and only by now start to cross the bridge to document standards.

However, we think that the standards and the commercial tools developed so far only partially offer the necessary prerequisites for multimedia document modeling as "next generation" multimedia applications extend the requirements given above by far: demand for *reusability* of the media including entire documents and parts of documents, modeling of *adaptation* to user specific needs and context-dependent, fine-grained re-usage of the multimedia material, and wide-spread use in the Internet.

Why do we consider these to be the new requirements to multimedia documents? As authoring of multimedia information is a very time consuming and costly task the reuse of material is definitely of high interest simply from an economical point of view. But reuse by means of "cut and paste" obviously can not be a solution, rather distinct and fine-grained reuse of multimedia content is highly demanded. Personalization and adaptation of information systems to personal needs and personal interests become more and more important (e.g., [Bul98]). The trend to offer a user the most suitable and narrowed down multimedia information can be seen in research prototypes from different research areas, e.g., a user adaptive diagram assistant [C-L98], an adaptive tutorial agent [SSW98], adaptive textbooks on the WWW [EBS97], personalized news paper [KBA93], personalized delivery of news [KLAV98], etc. Personalization and adaptation in consequence calls for the enhancement of multimedia content with metadata to allow for the targeted context-specific selection of multimedia content. Another new requirement to a document

model is its Internet-applicability, i.e., how it can cope with the demands of the heterogeneous environment of the Internet.

Within our project "Gallery of Cardiac Surgery" (Cardio-OP)[1], that aims at the development of an Internet-based and database-driven multimedia information system in the domain of cardiac surgery, we find a representative application that explicitly requires a model for multimedia material which can be extensively reused in different context. Based on a multimedia content repository, the system is going to serve as a common information and education base for its different types of users, physicians, medical lecturers, students, and patients, who are provided with multimedia information according to their user specific request to the multimedia information system, their different understanding of the selected subject, their location and technical infrastructure. For example, a high quality multimedia presentation dynamically composed during a lecture at the university campus should be available for students at home for revise although they do not need the high quality of videos or images at home. Therefore, either the multimedia material must be delivered to the student with a lower quality and lower bitrate or high data volume parts like video are replaced by "comparable" but less voluminous parts like a slide show.

To achieve this kind of functionality, a suitable multimedia document model that allows for flexible and context-dependent reuse of a multimedia document and parts of it is needed. On our way to such a "next generation" multimedia document model we first tracked down and identified the advanced requirements looking at upcoming advanced applications like Cardio-OP. When investigating the applicability of the existing models HTML, HyTime, MHEG, and SMIL we, unfortunately, find serious limitations and drawbacks. It is quite a challenge to push these limits and to define a multimedia document model providing modeling primitives that go beyond those found in existing models. An example for such a model is ZYX, presented in [BK99].

The implications of approaches trying to resolve the shortcomings of existing models are manyfold: There arises an urgent need for suitable authoring systems that support fine-grained reuse of multimedia content, adaptability of content to user needs and individual interest, and, as a direct consequence, the presentation-neutral representation of material, e.g., in a database. The latter is an analogue to the principle of data independence of applications well-known from database systems. What is known as data independence for "traditional" applications must be enhanced by presentation independence for multimedia applications. In addition, presentation-neutral representation of multimedia content directly impacts the design of presentation tools, since these have to "deliver" flexibility and adaptability to end users.

In this paper, we give a motivation towards the development of next generation multimedia document models like ZYX [BK99] by identifying advanced application requirements, analyzing existing document models showing the limits of current approaches, and calling for a concerted action on developing next generation authoring and management tools for advanced multimedia applications.

The remainder of the paper is organized as follows: Section 2 presents the new requirements for multimedia document models. Section 3 introduces the reader to the different models for multimedia documents we compare in this paper, HTML, SMIL, MHEG-5, HyTime, and ZYX. Section 4 presents the comparison of the models along the requirements identified in Section 2. The paper concludes with an reflection of the analysis and points the way to the future of multimedia document models.

2 Requirements to Next Generation Multimedia Document Models

In this section, we identify requirements to multimedia document models. These can be divided into traditional requirements, which we consider to be imperative for any multimedia document model, and advanced requirements, which we expect to be demanded more and more by future multimedia applications. The availability of a *temporal model*, a *spatial model*, as well as support for the modeling of *interaction* are traditional requirements while *reusability* of multimedia document content, *adaptation*

[1] Cardio-OP - Gallery of Cardiac Surgery - is partially funded by the German Ministry of Research and Education, grant number 08C58456. Our project partners are the University Hospital of Ulm, Dept. of Cardiac Surgery and Dept. of Cardiology, the University Hospital of Heidelberg, Dept. of Cardiac Surgery, an associated Rehabilitation Hospital, the publishers Barth-Verlag and dpunkt-Verlag, Heidelberg, FAW Ulm, and ENTEC GmbH, St. Augustin. For details see also URL www.informatik.uni-ulm.de/dbis/Cardio-OP/

to user specific needs, and *presentation-neutral representation* of multimedia document content are advanced requirements. Each of these requirements is motivated and illustrated in its different facettes in the following subsections. The requirements form a metric along which selected multimedia document models are analyzed in Section 4.

2.1 Temporal model

As the presentation of multimedia documents is time-dependent, one of the basic requirements to a multimedia document model is the modeling of the temporal course of the presentation. Thus, a temporal model must be provided to describe temporal dependencies between the media elements that a multimedia document comprises. We find three types of temporal models: *point-based* temporal models, *interval-based* temporal models and *event-based* temporal models.

In the point-based model the temporal extent of each media element in the multimedia document is modeled by *points in time*. These determine at which point in time on the time axis the presentation of a media element starts, and ends respectively. For any two points in time one of the relationships *before* ($<$), *after* ($>$), or *equals* ($=$) holds. This is a simple representation of time with a small number of temporal relationships.

Existing representations of temporal aspects in the context of multimedia presentations are mainly based on some or all of the 13 binary temporal relations between *time intervals* as defined by Allen [All83]. These models, however, do not support time intervals of unknown duration that occur, for instance, in the context of user interaction in multimedia presentations (e.g., Object Composition Petri Nets (OCPN) [LG93]). Therefore, enhanced interval-based temporal models have been proposed to handle open time intervals and indefinite interval relationships [DK95, HFK95, WR94].

In an event-based model of time, *events* determine the temporal course of the presentation. An event is connected to actions and when an event occurs, e.g., a video reaches a certain point in time, the corresponding actions, typically start and stop of the presentation of other media elements, is carried out.

Another way to specify temporal relations between media elements is by the use of *scripts* – programs written in a scripting language which can comprise temporal operations. If the scripting language forms a complete programming language, this mechanism allows for very complex and powerful specifications of temporal dependencies between media elements.

2.2 Spatial Model

If a presentation consists of visual media elements, not only the temporal synchronization of these elements is of interest but also their spatial positioning on the presentation media (e.g., a window). This positioning can be specified by the use of a spatial model. In general, three approaches to spatial models can be distinguished: *absolute positioning*, *directional relations*, and *topological relations*.

With absolute positioning the media element is placed on the presentation area at a fixed *absolute position* specified by a coordinate pair. To handle overlapping, a third value may be introduced by which the ordering of overlapping media elements is defined.

A more flexible way to define the spatial positioning of visual media elements is the specification of *directional* relations [PTSE95, PS94], like *north*, *north-west* etc. At a finer granularity, by introducing relations like *strong-north* and *weak-north* to specify overlapping, 169 different directional relations between two rectangles in 2D space can be distinguished [PTSE95].

Another way to define spatial relationships is by the use of topological relations [EF91]. Between any two continuous region objects, the following eight topological relations can be distinguished: *disjoint*, *meet*, *overlap*, *covers*, *covered-by*, *contains*, *inside*, and *equal*.

2.3 Interaction

A distinct feature of a multimedia document model is the ability to specify user interaction in order to let a user choose between different presentation paths. Multimedia documents without user interaction are not very interesting as the course of their presentation is exactly known in advance and, hence, could be recorded as a movie. For the modeling of user interaction, one can identify at least two basic types of interaction: *navigational interactions*, *design interactions*.

With navigational interactions a user can determine the flow of a multimedia presentation. An example is the selection of a link or an item from a menu to decide which presentation path is to be followed.

Design interactions influence the visual and audible layout of a presentation. Examples are the adjustment of speaker volume, fonts, scaling of images, and the like.

2.4 Reusability

As motivated in the introduction, reusability of multimedia content is a desired feature of a next generation multimedia document model. Reusability of document content can be characterized along three dimensions: the *granularity* of reuse, the *kind* of reuse, and the *selection and identification* of reusable components.

Granularity: The granularity of reuse determines *what* can be reused. Regarding multimedia document models, we can distinguish at least three levels of granularity of reusable components: reuse of complete multimedia *documents*, reuse of *fragments* of multimedia documents like single scenes or chapters, and reuse of individual atomic *media elements* such as a video or audio.

Kind of re-usage: For all three levels of granularity we distinguish between different ways of *how* to reuse material for the composition of new documents: *identical re-usage*, i.e., the components are reused including all temporal, spatial, design and interaction relationships and constraints as originally specified by the author(s), and *structural re-usage*, i.e., we separate the layout from the structure of components and reuse only the structural parts.

Selection and identification: Before we can reuse components we have to *identify* and *select* them within an information system. This calls for metadata and a mechanism for classifying, indexing, and querying components. Hence, a document model should provide support for annotation of reusable components with metadata.

2.5 Adaptation

Presentation of multimedia documents preferably depends on the user context and hence, the multimedia presentation needs to be adapted to this user context. But it is also of interest whether all possible adaptation alternatives are to be known and modeled at authoring time of a multimedia document or if they are left for evaluation at the actual presentation time.

Parameters of adaptability: For the user context, we distinguish between *adaptation to personal interest* and *adaptation to technical infrastructure*. Consider a professor on campus who is interested to see in-depth multimedia material on coronary artery bypass grafting, and an undergraduate student at home who needs to get only an abstraction of the same material. In the example the presentation needs to be adapted to the personal interest, here identified by personal interest "coronary artery bypass grafting" and professional level "professor" and "student". In addition to this kind of semantic adaptation of multimedia documents, the multimedia presentation can be adapted according to the technical capabilities of the environment a user is working in, i.e., "on campus", "at home". The professor may run a high quality presentation on the university campus providing excellent network bandwidth and computer power, whereas the student can view the presentation at home where he does not have the same excellent technical prerequisites. A document model supports these types of adaptation if it can support the modeling of user-specific and system-specific parameters as "input parameters" for adaptation sufficiently.

Definition of presentation alternatives: Depending on *when* the different "alternatives" are defined that can be exploited for adaptation, we distinguish between *static adaptation* and *dynamic adaptation*. With static adaptation the adaptable alternatives must be known and included in the document at authoring time. Whereas for dynamic adaptation the available alternatives are determined due to the specific context at presentation time. One could therefore say that models that allow static and/or dynamic adaptation allows for "early and/or late adaptation binding".

2.6 Presentation-neutral Representation

Reuse of multimedia content in different context does not mean that the material is presented always identically. Rather reuse of content may require structural reuse of material and assignment of different visual and audible layout according to the context. In addition, advanced distributed multimedia applications often face a heterogeneous environment with regard to operating systems and hardware platforms. It is desirable that the multimedia material of such an application can be presented within this heterogeneous environment with minimal implementation effort. Thus, it makes perfect sense to try to reuse existing presentation software, e.g., HTML browsers, MHEG engines, on these systems.

As a consequence, the multimedia material has to be modeled in a *presentation-neutral* way, i.e., independent of the actual realization (layout) of a presentation. This is a challenging problem as it calls for automatic conversion of the multimedia document model used for the presentation-neutral description of multimedia content into the multimedia document model used for presentation of the multimedia content. In general, two major characteristics influence the convertability between multimedia document models: *multimedia functionality* and *semantic level* of a model [RvOB97].

Multimedia functionality: The multimedia functionality of a multimedia document model describes the expressiveness of its modeling primitives. In a conversion process, this means that if the target document model does not offer an equivalent multimedia functionality as offered by the source model, the conversion will be lossy.

Semantic level: The semantic level of a multimedia document model plays an important role for the automatic conversion for presentation. If the target document model of such a conversion provides a semantic description of multimedia content on a high level, i.e., rather description of structure than description of presentation, higher than the source document model, the conversion requires the analysis of the document specified in the source model and the derivation of its semantics for its encoding in the target model. In general, this requires knowledge about the multimedia content that often only the author will have. In these cases, automatic conversion will not be possible. However, the automatic conversion of a multimedia document represented on a high level of semantics into a model based on a comprehensive set of low level semantic constructs can be performed much easier. In order to avoid the problems of automatic conversion, the presentation-neutral representation of multimedia content should – besides the coverage of rich multimedia functionality – take place on a high level of semantics.

3 Existing Multimedia Document Models

In this section, we briefly present the most important and relevant existing standards and data models for multimedia documents. We give an introduction to the existing document models HTML, SMIL, MHEG-5, HyTime, and also to ZyX, an example for a model offering more advanced modeling primitives. These models will be analyzed and compared along the requirements of the previous section in Section 4.

3.1 HTML

The Hypertext Markup Language (HTML) [RLJ98] is based on SGML [ISO86] and defines a syntax to enrich text pages with structural information using SGML *elements*. For instance, elements can be inserted into the text to organize it into paragraphs, to mark headings of different levels, to define tables, and to define quotations. Furthermore, it is possible to include various kinds of objects like media elements (e.g, images, videos and audio tracks), Java applets, ActiveX components, and scripts. In addition to that, HTML allows for the definition of *hyperlinks* between documents. These hyperlinks are means to define interactions, i.e., an interaction (e.g., a mouse click) with the *link anchor* results in the presentation of the document specified by the *link target*. Scripts, applets, and ActiveX components included with a document are executed at presentation time by the presentation environment, the so-called *HTML browser software*. However, the HTML standard does neither define syntax nor semantics of the scripting languages, so presentation behaviour of a HTML page that includes scripts depends on the employed browser software.

There are efforts of the large HTML browser software vendors Netscape and Microsoft to allow for the manipulation of the structure, layout, and content of a HTML document with scripting languages. Thus,

scripts can dynamically manipulate HTML documents, a technique which is also called *Dynamic HTML (DHTML)*. The price for this increased flexibility is that portability problems arise due to differences between the scripting languages employed by Microsoft and Netscape.

3.2 SMIL

The Synchronized Multimedia Integration Language (SMIL) [HBB+98] is a W3C standard which aims at synchronized multimedia presentations on the web. A SMIL document provides synchronization of continuous media elements and constitutes an integrated presentation. SMIL is defined by an XML DTD [BPSM98] and, hence, the language can be understood as a set of element definitions specified in terms of XML. SMIL defines *schedule elements* to describe temporal synchronization between media elements. Furthermore, the spatial layout of the media elements can be defined. SMIL also allows to specify links between documents or parts of documents which are equivalent to HTML links. An interesting feature of SMIL is the *switch* element which is a simple means for modeling alternatives in the course and quality of a presentation. With the help of switch elements, an author can specify different presentation alternatives among which one is chosen at presentation time due to external parameters.

3.3 MHEG-5 and MHEG-6

MHEG-5 [ISO95, JR95] is an adaptation of the MHEG-1 Standard [MBE95] to the needs of video-on-demand and kiosk applications for set-top-boxes and low-end PC. MHEG-5 encodes applications in their final form and aims at an efficient realization of MHEG-1 attracting the interest of telecommunication and entertainment industry in this standard. MHEG-5 provides an object-oriented data model for multimedia documents. The standard defines a hierarchy of *MHEG-5 classes*. This hierarchy comprises classes for various uses. For example, there are classes that represent media elements like videos and audios, classes that represent interaction elements like buttons, and even classes that provide variable functionality of programming languages. Classes possess attributes, can perform actions (which closely resemble methods in object-oriented programming languages), and fire events. An MHEG-5 document is a collection of instances of these classes organized in *scenes* which are the main structural primitives. A scene corresponds to a "page" on a screen and, hence, only one scene can be presented at a time. In addition to that, each MHEG-5 document features one instance of the class `Application` defining the entry point for document's presentation. Moreover, this application object can contain objects which are global to every scene. The presentation behaviour of an MHEG-5 document is defined by the means of *links* which resemble event-condition-action rules.

MHEG-6 [ISO96, Hof96] is an extension of MHEG-5 that introduces an interface between an MHEG-5 engine and a Java Virtual Machine. With MHEG-6, it is possible to include Java programs into an MHEG-5 document. Such a program has access to the objects of the document and, hence, can influence its presentation behaviour.

3.4 HyTime

HyTime [ISO92, DD94, NKN91] is a standard which allows for the description of the structure of multimedia documents. Based on SGML [ISO86], HyTime provides a well-defined set of primitives which allows for the interlinking of media objects without specifying the encoding of the media objects. The primitives provided by HyTime are offered by means of *architectural forms* and are organized in terms of modules. Architectural Forms (AF) are HyTime elements with pre-defined multimedia semantics and attributes. An AF can be used in any SGML DTD by extending an SGML element type by an attribute `HyTime` bearing the name of the AF to be used. In that way, the element type inherits the semantics and attributes of the AF.

The modules of HyTime Base-Module, which defines the basic concepts of HyTime, the Location-Address-Module, which implements the powerful construct of *Locators* providing an abstract mechanism for addressing external document objects, the Hyperlink-Module, which implements the concept of links, the Finite-Coordinate-Space-Module, which provides means for the synchronized presentation of media objects based on n-dimensional *coordinate spaces*, the Event-Projection-Module, which allows to transform event schedules defining the temporal execution of a presentation, and the Object-Modification-Module, which allows to transform presentation objects, e.g., fading.

3.5 ZyX

The ZyX multimedia document model [BK99] has been developed by our group in the context of the Cardio-OP project which aims at the development of a database-driven multimedia information system with special needs for reusability, adaptation, interaction, and presentation-neutral description of multimedia content. ZyX describes complete or fragments of multimedia documents by the means of a tree (for an illustration, see Figure 1). The nodes of the tree are called *presentation elements*. Each presentation has got a *binding point* associated with it. Such a binding point can be bound to one *variable* of another presentation element, thus creating the edges of the tree. The presentation elements are the generic elements of the model. They can represent *atomic media elements* (e.g., videos, images and text) or more complex compositions of media elements. Another group of presentation elements combine presentation elements with certain semantics, the *operator elements*. There are operator elements that allow for temporal synchronization, definition of interaction, adaptation, and for the spatial, audible, and visible layout (the so-called *projector elements*) of the document.

It is possible to delay the process of variable binding by leaving variables unbound. This allows for the definition of *templates* which can be customized to a specific problem at a later point in time. Furthermore, a tree can be encapsulated by a *complex media element* which can then be used in other trees (see Figure 2) like any other presentation element. Unbound variables of an encapsulated tree are exported by the complex media element allowing for the encapsulation of templates. Thus, a complex media element is somehow a black box view of a ZyX tree.

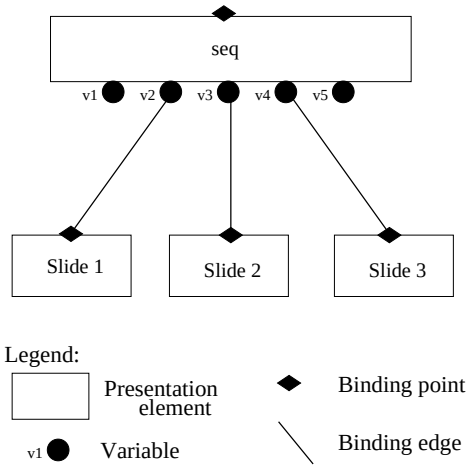

Figure 1: A sample ZyX tree in graphical notation

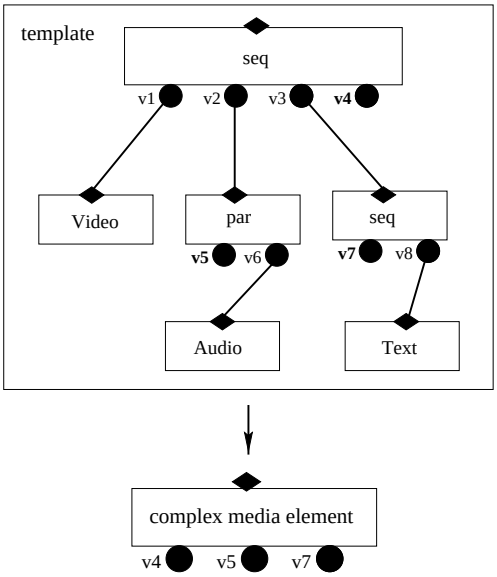

Figure 2: Template encapsulated by a complex media element in graphical notation

4 Analysis

In this section, we analyze how the multimedia document models introduced in the previous section fulfill the requirements outlined in Section 2. First, we investigate the document models HTML, MHEG-5, HyTime, SMIL, and ZyX with respect to the traditional requirements, i.e., their temporal and spatial model, and their interaction modeling. Then, we examine how these models behave concerning the advanced requirements reusability, adaptation, and presentation-neutral description for document content. Figure 4 illustrates the summary of this analysis.

4.1 Temporal Model

HTML: As HTML has been developed for Hypertext, the standard itself does not offer constructs to specify temporal synchronization between the media elements included in a HTML document. However, by the use of DHTML, the temporal course of a presentation can be programmed in a scripting language. Recently, there has been a proposal by Microsoft for adding temporal synchronization support to HTML called HTML+TIME [SYS98]. This approach integrates HTML with the temporal modeling mechanisms introduced by SMIL but it does not seem to be very mature, yet.

MHEG-5: MHEG-5 specifies the temporal course of a presentation by means of its link concept. Presentable MHEG classes (e.g., the video and audio classes) define a variety of events relating to time. These events can be associated via links with actions which are performed when the corresponding event occurs. Thus, the temporal model of MHEG-5 is event-based.

HyTime: In HyTime, the coordinated presentation of media elements can be specified using the Finite-Coordinate-Space module. This module provides a means to define n-dimensional coordinate spaces which can include time dimensions. Into such a coordinate space, media elements can be placed. This placement is called *event* and exactly defines the point in the coordinate space where the media element associated with the event will be presented. Hence, the temporal model of HyTime is point-based. Several events can be grouped to an *event schedule* which describes the course of the presentation of a HyTime document.

SMIL: SMIL follows an interval-based approach to temporal synchronization. Each media element has an associated presentation interval. These intervals can be coordinated by the use of schedule elements. In general, SMIL defines two kinds of schedule elements. On the one hand, there is the *parallel element* which defines the parallel presentation of n intervals. Using attributes, a more detailed definition of a parallel presentation is possible. For instance, time delays, lipsync synchronization, and loops can be specified. On the other hand, SMIL provides the *sequential* element which allows for the sequential presentation of n intervals. Again, a more detailed specification of this presentation is possible with the help of element attributes. The different schedule elements can be nested, thus allowing for the modeling of complex temporal relations.

ZYX: The temporal model of ZYX is closely related to the temporal model of SMIL. ZYX defines the temporal operator elements *seq* and *par* which resemble the parallel and sequential elements of SMIL. In contrast to SMIL, loops and time delays are not specified using attributes. Instead, these are handled by own temporal operator elements, the *loop* operator element and the *delay* temporal operator element. Due to its close resemblance to SMIL, the ZYX temporal model must be considered as interval-based.

4.2 Spatial Model

HTML: The control of the spatial layout of the media elements included in a HTML document is very limited. However, the concept of *framesets* allows to partition the presentation area of an HTML document (i.e., the HTML browser window) into rectangular regions, so-called *frames*. In such a frame, another HTML document can be displayed which itself can define further frames. This allows for frame nesting. As frames are specified by their size and position this constitutes a kind of absolute positioning. Exact positioning of media elements is possible by the use of DHTML. Scripts can set and modify the coordinates at which a media element is presented in the browser window. Hence, (D)HTML offers absolute positioning.

MHEG-5: Each MHEG-5 class that represents visual media elements provides attributes defining the coordinates of the presentation area at which the visual media element has to be presented. By the use of the link concept, these coordinates can be set and changed as the result of events. This is a kind of absolute positioning.

HyTime: As mentioned above, the Finite-Coordinate-Space module provides means to specify the course of the presentation of a HyTime document by the means of event schedules referring to n-dimensional coordinate spaces. Such a coordinate space can include, besides a temporal dimension,

one or more spatial dimensions. Thus, an event not only describes the point in time at which the associated media element will be presented but also the spatial position. Therefore, HyTime allows for absolute positioning.

SMIL: SMIL provides a mechanism to allow for the absolute spatial positioning of media elements. In the head of a SMIL document, rectangular regions of the presentation area can be specified, called *channels*. Each channel is defined by its position, its size, and a value which is used to define the order of overlapping channels. Each media element in the document body can reference a channel thereby specifying its spatial position on the presentation area.

ZyX: Spatial layout in ZyX is defined by the use of *spatial* projector elements. A spatial projector element defines the rectangular region of the presentation area in which the subtree below the spatial projector element is presented. Like channels in SMIL, such a region is defined by its position, size, and a value to resolve overlapping of regions. Spatial projector elements can be nested. A spatial projector element p in the subtree under spatial projector element o is seen in the context of o and not the entire presentation area. All in all, ZyX employs absolute positioning to specify spatial layout of a document.

4.3 Interaction

HTML: HTML provides the concept of links which allows for navigational interaction. Moreover, HTML allows for the definition of data entry forms which consist of different controls like buttons and text-input fields. The results of an interaction with a form cannot be specified using HTML itself. This is left to CGI scripts running at a web server. However, if the employed browser software supports DHTML, more sophisticated interactions like design interactions can be programmed using scripts.

MHEG-5: MHEG-5 provides a small set of basic interaction classes for the modeling of user interaction. MHEG-5 separates the element that initiates an interaction from the effect of an interaction. By the definition of links, the interaction with an interaction element such as a button can trigger an action resulting in a navigational or in a design interaction. Since MHEG-6 allows for the integration of Java programs, it can support additional, sophisticated user interactions.

HyTime: As explained above, the Finite-Coordinate-Space module of HyTime provides mechanisms to define the spatial and temporal coordination of a presentation. This is done by event schedules which require to know all spatial and temporal positions of media objects in advance. This excludes ad-hoc navigational interaction by the user. HyTime does not support design interactions as the primitives provided by the Object-Modification-Module and the Event-Projection-Module do not include *interaction* semantics.

SMIL: The concept of links in SMIL provides for navigational interaction. But no support is given for the specification of design interactions.

ZyX: The requirement to support the modeling of interactive multimedia presentations is met by ZyX's interaction elements. There are two types of interaction elements, *navigational* interaction elements and *design* interaction elements. Examples for navigational interaction elements are the *link* element that allows to specify hypertext structure as in SMIL or HTML and the *menu* element with which one can interactively follow one path out of a set of presentations paths. The design interaction elements are interactive versions of the projector elements. For example, for the typographic projector that allows to specify font, size and style of a text, the *interactive typographic projector element* specifies that these settings can be carried out interactively when the document is presented.

4.4 Reusability

HTML: HTML allows to reference whole documents and single media elements via *uniform resource locators (URL)*. However, it is not possible to reference just a fragment of a document. Thus, reusability is only supported on the highest and lowest level of granularity as identified in Section 2.

As there is no clear distinction between structure and layout of an HTML document, hence, reuse can only be identical. Although HTML 4.0 tries to separate structure from the layout of a document in a more rigid way and promotes the use of external cascading style sheets, it is still possible to mix layout and structure for backward compatibility reasons.

In order to support classification and identification of documents, HTML allows for the specification meta attributes by means of attribute-value pairs in the head of a document.

MHEG-5: Considering the granularity of reuse, it is important to notice that MHEG-5 structures the media elements of an application into *groups* which can be addressed globally. Hence, groups constitute the units which can be reused. As an application object is a group, it is possible to reuse entire MHEG-5 documents within an MHEG-5 document. Likewise, scenes are MHEG-5 groups and, hence, could be reused in principal. Since scenes can refer to objects global to an MHEG-5 document which are contained in the application object, it is not possible to reuse scenes which depend on such global objects. Only fully independent and isolated scenes could be candidates for re-usage in other documents. Therefore, there is no general support for reusability at the level of document fragments. Regarding the reusability at the level of mere media elements (like videos, audios), it is important to know that a media element must be associated to exactly one group and is addressed through this group. Thus, it is not possible to use one and the same media element in two different scenes. However, as media elements do not have to include the underlying data but also just can refer to their data, groups can share at least the data of media elements.

Since MHEG-5 aims less at modeling the structure of a multimedia application but at representing its final presentation form, which includes the layout, groups can only be reused identically.

The identification and selection of groups to be reused is a serious problem in MHEG-5 as no information can be assigned to MHEG objects. One can use neither annotations, keywords, or any other kind of metadata for classification of and search for media objects, nor semantically useful names for the identification of parts to be reused.

HyTime: HyTime allows for reusability on all levels of granularity as identified in the requirements section. As HyTime is built on SGML, single media elements and complete documents can be referenced as entities and therefore be reused. Moreover, the Location-Address-Module provides powerful mechanisms to locate and address fragments of HyTime documents. Using *locators*, parts of a HyTime document can be referenced by name, position, or even by the use of a powerful query language.

As any SGML DTD can be made HyTime-compliant, HyTime documents describe rather the structure of a document than its presentation semantics. Thus, reuse in HyTime is semantic reuse.

Moreover, because HyTime is independent of a DTD, a DTD can be provided with support for classification of (parts of) documents, e.g., by the use of attribute-value pairs. Hence, HyTime offers support for classification and identification of reusable components.

SMIL: As SMIL can reference complete documents and single media elements by the use of URL, SMIL allows for reuse at the according levels of granularity as defined in Section 2. However, SMIL does not support the reuse of fragments of documents.

SMIL separates layout specifications, which have to go into the head of the document, from the structural specifications given in the body. But as both kinds of specifications are closely interrelated, SMIL provides only for identical reuse.

Like HTML, SMIL allows to define meta-attributes within the head element of a document. Such meta-attributes can be used to classify and retrieve documents providing support for selection and identification.

ZyX: The ZyX document model has been designed with all levels of granularity of reuse in mind. To support reusability of media elements, atomic media elements are provided which can be reused in any ZyX specification. Likewise, complex media elements which encapsulate specifications can be reused in any other specification. As the encapsulated specifications can smoothly range from small logical parts of a document to entire documents, ZyX supports reuse both on the level of entire documents and fine-grained document fragments. Moreover, the ability to encapsulate templates in complex media elements provides for the reuse of document templates.

The ability to delay the process of variable binding, especially the binding of projector variables, allows for the clear separation of the presentation elements building the structure of a document and the

projector elements determining its layout. This allows for structural reuse of ZYX specifications. As ZYX complex media elements may include project elements defining visual and audible layout this provides for identical reuse of components.

Concerning selection and identification of reusable elements, ZYX allows media elements, either complex or atomic, to be annotated with key-value pairs.

4.5 Adaptation

HTML: Since HTML does not offer any mechanism to specify adaptation of a document to user interest or to technical infrastructure, we consider only DHMTL here. DHTML offers to dynamically manipulate the structure and content of HTML documents. Therefore, adaptation to user interest or technical infrastructure can be implemented by the use of scripts. In a first step, such a script has to determine the user or system profile, for example by a database query. In the second step, the script has to change the structure of the HTML document according to the profile. As the author must code and thus know at authoring time all adaptation alternatives inside scripts, this kind of adaptation must be considered as static.

MHEG-5: MHEG-5 defines classes for variables whose contents can be tested. Hence, variables can be used to choose between different branches of a presentation. Thus, a profile defining user interest and technical infrastructure could be modeled using variables. However, the problem is how such a profile is set. MHEG-5 allows to set variables only from within a document. User-specific adaptation would require to make the determination of the profile a part of the MHEG-5 document. In MHEG-6, the MHEG engine could call a Java program which retrieves the actual values for a given profile and then sets the variables of the document. So, with the use of MHEG-6, adaptation of a presentation to user interest or technical infrastructure is possible. Since all adaptation alternatives must be specified within a document at authoring time, this is static adaptation.

HyTime: Since HyTime can be used with any concrete DTD, it is always possible to define specific attributes with elements of a DTD that characterize (parts of) documents or media elements in terms of user interest or technical properties like bandwidth needed, resolution or frame rate required. It is also possible to check for values of such element attributes by using the Query-Locator provided by the Location-Address-Module. But all the results of such queries checking attribute values are fully determined by the concrete document content and cannot be modified by external parameters like those in a user or system profile. Hence, it is not possible to adapt a HyTime document according to external parameters like a profile.

SMIL: SMIL offers the *switch* element to model alternative presentation variants. Using this element, different adaptation alternatives can be specified inside the document at authoring time. Thus, the switch element allows for static adaptation. The selection of the alternatives is guided by simple predicates which include parameters set outside the SMIL document. These parameters are predefined by the standard and describe mainly technical features like the available bandwidth. This allows to adapt a SMIL document to technical infrastructure.

ZYX: As mentioned above, each media element of a ZYX document can be annotated with a set of key-value pairs that describes its content. In addition to that a user profile, also key-value pairs, can be defined to capture values that describe a user's topics of interest, presentation system environment, network connection characteristics and the like. The ZYX model offers operator elements to support adaptation to a user's profile by means of *switch elements* and the *query elements*.

Like in SMIL, a switch element allows to specify different presentation alternatives for a part of the document allowing for static adaptation. One of the alternatives is selected corresponding to the user profile. In contrast to SMIL, the scope of a switch statement is not limited to predefined parameters. A switch element is used if all adaptation alternatives are known to the author of a document. In order to allow for dynamic adaptation, the *query* element is provided. This element is a placeholder for a media element or fragment which is described by the means of a query. The query is represented by a set of key-value pairs. When the document is selected for presentation the query element is evaluated

and replaced by the complex or atomic media element with best matching the set of key-value pairs with regard to the user profile. Thus, ZYX allows for adaptation to user interest and system structure.

4.6 Presentation-neutral Representation

Figure 3 shows the relationships between various formats and models with respect to their support for multimedia and for presentation-neutral representation.

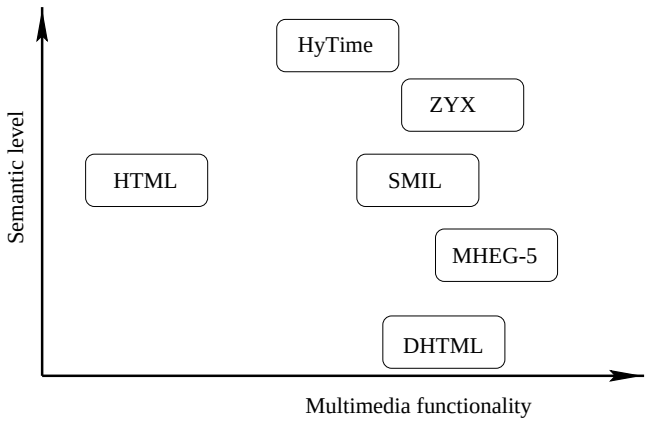

Figure 3: Presentation-neutral representation and multimedia support of the different formats and models

HTML: As HTML does not clearly separate the layout of a document from its structure, the semantic level of a HTML document description is not as high as HyTime though it is comparable to SMIL. However, HTML offers only an extremely limited multimedia functionality (even simple temporal synchronization is not possible). To offer more multimedia functionality, DHTML must be employed. Since DHMTL scripts must imperatively implement multimedia functionality, their use extremely reduces the semantic level of a document description. Furthermore DHTML introduces portability problems between browser vendors. Thus, neither HTML nor DHMTL are well suited for presentation-neutral representation of multimedia document content.

MHEG-5: MHEG-5 primarily aims at a detailed and platform independent description of a presentation, i.e., the layout, of a document. To achieve this goal, the standard provides MHEG-5 with a rich multimedia functionality. However, the description of the structure of an MHEG-5 document is very poor. Hence, the level of semantic modeling is very low and if compared to the semantics of the structure of multimedia documents MHEG-5 might be viewed as a "multimedia-assembler". Hence, MHEG-5 cannot support presentation-neutral representation of multimedia documents.

HyTime: Since HyTime mainly specifies the structure and semantics of a multimedia document it is quite well-suited for presentation-neutral representation of multimedia documents. HyTime offers specification of document content at a high semantic level though it lacks multimedia functionality (especially in the area of interaction).

SMIL: In contrast to HyTime documents, a SMIL document describes in detail the presentation of the document but less detailed the structure of the document. However, SMIL offers more multimedia functionality than HyTime. Compared to MHEG-5, the description of SMIL documents takes place on a higher level of semantics though lacking the multimedia functionality of MHEG-5. Hence, SMIL ranks between HyTime and MHEG-5 with respect to its support for presentation-neutral representation.

ZYX: As it is possible to separate structure and layout of a document due to the ability to delay the process of variable binding and to encapsulate templates in complex media elements, the semantic level of a document description is quite high and thus suited for presentation-neutral representation of multimedia document content. The amount of multimedia functionality offered by ZYX exceeds SMIL but ranks below MHEG-5.

	HTML	DHTML	SMIL	MHEG-5	HyTime	ZYX
Temporal Model	-	script	interval-based	event-based	point-based	interval-based
Spatial Model	absolute positioning	absolute positioning	absolute positioning	absolute positioning	absolute positioning	absolute positioning
Interaction						
Navigational	+	+	+	+	-	+
Design	-	+	-	+	-	+
Reusability						
Granularity						
Media Elements	+	+	+	+	+	+
Fragments	-	-	-	-	+	+
Documents	+	+	+	+	+	+
Kind of Reusage						
Identical	+	+	+	+	-	+
Structural	-	-	-	-	+	+
Identification/Selection	+	+	+	-	+	+
Adaptation						
Parameters of Adaptability						
User Interest	-	+	-	MHEG-6	-	+
Technical Infrastructure	-	+	+	MHEG-6	-	+
Definition of Alternatives						
Static	-	+	+	MHEG-6	-	+
Dynamic	-	-	-	-	-	+
Presentation-neutral Representation						
Multimedia Functionality	very low	high	medium	very high	low	high
Semantic Level	medium	very low	medium	low	very high	high

Figure 4: Summary of the support of the requirements by (D)HTML, SMIL, MHEG-5, HyTime, and ZYX (+ support, - no support)

4.7 Summary

Summarizing (see also Figure 4), we can say that none of the examined document models HTML, MHEG-5, HyTime, and SMIL offers sufficient support for all requirements arising from advanced multimedia applications. HTML can hardly be characterized as a multimedia document model because it lacks support for even the most basic multimedia requirement, a temporal model. Though HTML can become a quite powerful multimedia document model by the extension to DHTML, it still lacks support for reuse at all levels of granularity and suffers from a low semantic level of content description which leaves DHTML unsuitable for presentation-neutral description of multimedia content. This is also the case with MHEG-5. Although MHEG-5 offers a high multimedia functionality, it mainly describes the presentation and not the structure of a multimedia document and, therefore, cannot be employed for presentation-neutral modeling of multimedia document content. Furthermore, reuse at the level of fragments is severely hampered due to the unflexible scene-based document structure.

Powerful support for reuse is the strength of HyTime. Moreover, HyTime describes document content at a very high semantical level and, thus, is perfectly suited for presentation-neutral modeling of document content. However, the lacking capability of interaction modeling and modeling of adaption is a serious drawback. In contrast to HyTime, SMIL offers the modeling of static adaptation to technical infrastructure and navigational interaction. Furthermore, the semantic level of a SMIL document description ranks between MHEG-5 and HyTime and, hence, is quite well suited for the presentation neutral description of multimedia document content. However, reuse at the level of fragments is not possible as

is the modeling of design interactions.

Since ZYX has been designed with the fulfilment of the advanced requirements in mind, it offers reuse on all three levels of granularity, static and dynamic adaptation to user specific needs, a quite high semantic level of document description, and presentation-neutral representation of multimedia content. Regarding the traditional requirements, enough multimedia functionality has been provided to allow for interesting multimedia presentations including design interactions.

5 Conclusion and Future Work

Driven by our advanced multimedia information system application Cardio-OP, we first have identified a new set of requirements for multimedia document models: *reusability of multimedia content*, *adaptation of multimedia content to user needs and interests*, and *presentation-neutral description* of the structure and content of multimedia documents. These requirements complement the more traditional requirements for multimedia document models, i.e., temporal, spatial, and interaction modeling, well known so far. We then have presented an analysis of the relevant standard formats and models, i.e., HTML, SMIL, MHEG-5, HyTime, including the ZYX model which has been designed to meet the advanced requirements. We have presented the capabilities and identified the limitations of these models. The shortcomings of standards call for a new initiative for next generation multimedia document models. As illustrated by ZYX [BK99] it is very well possible to push the limits of existing approaches and to meet the new requirements.

We would like to point out that the implications of our analysis and of approaches trying to resolve the shortcomings of existing models are significant: There arises an urgent need for appropriate authoring tools that support fine-grained reuse of multimedia content, adaptability of content to user needs and individual interest, and, as a direct consequence, the presentation-neutral representation of material, e.g., in a database. When developing the multimedia content repository of Cardio-OP based on ZYX we made this painful experience. Our group has already developed a DataBlade module for the object-relational database system Informix Dynamic Server / Universal Data Option capable of managing ZYX documents and fragments [BKW99]. We currently develop an authoring tool and a presentation engine for ZYX, since presentation-neutral representation of multimedia content as well as adaptation support directly impacts the design of authoring and presentation tools.

References

[All83] J. F. Allen. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11):832–843, November 1983.

[BK99] S. Boll and W. Klas. ZYX — A Semantic Model for Multimedia Documents and Presentations. In *To be published in: Proceedings of the 8th IFIP Conference on Data Semantics (DS-8): "Semantic Issues in Multimedia Systems"*. Kluwer Academic Publishers, Rotorua, New Zealand, 5-8 January 1999.

[BKW99] S. Boll, W. Klas, and U. Westermann. Exploiting OR-DBMS Technology to Implement the ZYX Data Model for Multimedia Documents and Presentations. In A. Buchmann, editor, *Submitted to: Datenbanksysteme in Büro, Technik und Wissenschaft (BTW)*. GI-Fachtagung, March 1999.

[BPSM98] T. Bray, J. Paoli, and C. M. Sperberg-McQueen. *Extensible Markup Language (XML) 1.0 – W3C Recommendation 10-February-1998*. W3C, URL: http://www.w3.org/TR/1998/REC-xml-19980210, Februar 1998.

[Bul98] D. C. A. Bulterman. User-centered Abstractions for Adaptive Hypermedia Presentations. In *Proc. of the 6th ACM Multimedia Conference*, Bristol, UK, September 1998.

[C-L98] C-LAB — Research Institute of the University of Paderborn and Siemens AG. IDIAS - An Intelligent Diagram Assistant, 1998. URL http://www.c-lab.de/ ucmm/idias/root.html.

[DD94] S. DeRose and D. G. Durand. *Making Hypermedia Work: A User's Guide to HyTime*. Kluwer Academic Publishers, Dordrecht, 1994.

[DK95] A. Duda and C. Keramane. Structured temporal composition of multimedia data. In *Proc. IEEE International Workshop on Multimedia- Database-Management Systems*, Blue Mountain Lake, August 1995.

[EBS97] J. Eklund, P. Brusilovsky, and E. Schwarz. Adaptive textbooks on the www. In H. Ashman, P. Thistewaite, R. Debreceny, and A. Ellis, editors, *Proceedings of AUSWEB97, The Third Australian Conference on the World Wide Web, Queensland, Australia*, pages 186—192. Southern Cross University Press, July, 5–9 1997.

[EF91] M. J. Egenhofer and R. Franzosa. Point-Set Topological Spatial Relations. *Int. Journal of Geographic Information Systems*, 5(2), March 1991.

[HBB+98] P. Hoschka, S. Bugaj, D. Bulterman, et al. *Synchronized Multimedia Integration Language – W3C Working Draft 2-February-98.* W3C, URL: http://www.w3.org/TR/1998/WD-smil-0202, Februar 1998.

[HFK95] N. Hirzalla, B. Falchuk, and A. Karmouch. A Temporal Model for Interactive Multimedia Scenarios. *IEEE Multimedia*, 2(3):24–31, Fall 1995.

[Hof96] P. Hofmann. MHEG-5 and MHEG-6: Multimedia Standards for Minimal Resource Systems. Technical Report, Technische Universität Berlin, April 1996.

[ISO86] ISO. *Information processing - Text and Office Systems - Standard Generalized Markup Language (SGML)*, 1986. ISO-IS 8879.

[ISO92] ISO/IEC. *Information Technology - Hypermedia/Time-based Structuring Language (HyTime)*, 1992. ISO/IEC IS 10744.

[ISO95] ISO/IEC JTC1/SC29/WG12. *Information Technology – Coding of Multimedia and Hypermedia Information – Part 5: Support for Base-Level Interactive Applications, ISO/IEC IS 13522-5.* ISO/IEC, 1995.

[ISO96] ISO/IEC JTC1/SC29/WG12. *Information Technology – Coding of Multimedia and Hypermedia Information – Part 6: Support for Enhanced Interactive Applications, ISO/IEC IS 13522-6.* ISO/IEC, 1996.

[JR95] R. Joseph and J. Rosengren. MHEG-5: An Overview. Technical Report, GMD-FOKUS, Berlin, URL://www.fokus.gmd.de/ovma/mug/archives/doc/mheg-reader/rd1206.html, December 1995.

[KBA93] T. Kamba, K. Bharat, and M. C. Albers. The Krakatoa Chronicle - An Interactive, Personalized Newspaper on the Web. page http://www.w3.org/Conferences/WWW4/Papers/93/, 1993.

[KLAV98] W. Klippgen, T. D. C. Little, G. Ahanger, and D. Venkatesh. The Use of Metadata for the Rendering of Personalized Video Delivery. In *[SK98]*, New York, 1998. McGraw-Hill.

[LG93] T. D. C. Little and A. Ghafoor. Interval-Based Conceptual Models for Time-Dependent Multimedia Data. *IEEE Transactions on Knowledge and Data Engineering*, 5(4), August 1993.

[MBE95] T. Meyer-Boudnik and W. Effelsberg. MHEG Explained. *IEEE Multimedia*, 2(1), Spring 1995.

[NKN91] S. R. Newcomb, N. A. Kipp, and V. T. Newcomb. "HyTime" – The Hypermedia/Time-Based Document Structuring Language. *Communications of the ACM*, 34(11), November 1991.

[PS94] D. Papadias and T. Sellis. Qualitative Representation of Spatial Knowledge in Two-Dimensional Space. *VLDB Journal*, 3(4), October 1994.

[PTSE95] D. Papadias, Y. Theodoridis, T. Sellis, and M. J. Egenhofer. Topological Relations in the World of Minimum Bounding Rectangles: A Study with R-Trees. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, San Jose, May 1995.

[RLJ98] D. Raggett, A. Le Hors, and I. Jacobs. *HTML 4.0 Specification – W3C Recommendation, revised on 24-April-1998.* W3C, URL: http://www.w3.org/TR/1998/REC-html40-19980424, April 1998.

[RvOB97] L. Rutledge, J. van Ossenbruggen, and D. C. A. Bulterman. A Framework for Generating Adaptable Hypermedia Documents. In *Proc. ACM Multimedia Conference*, Seattle, November 1997.

[SK98] A. Sheth and W. Klas. *Multimedia Data Management - Using Metadata to Integrate and Apply Digital Media.* McGraw-Hill, New York, 1998.

[SSW98] V. Schöch, M. Specht, and G. Weber. ADI — An Empirical Evaluation of a Tutorial Agent. In T. Ottmann and I. Tomek, editors, *Proceedings of the ED-Media and ED-TELECOM 1998, Freiburg, Germany.* Association for the Advancement of Computing in Education, June 1998. URL http://apsymac33.uni-trier.de:8080/ADI.html.

[SYS98] D. Schmitz, J. Yu, and P. Satangeli. *Timed Interactive Multimedia Extensions for HTML (HTML+TIME).* W3C, URL: http://www.w3.org/TR/1998/NOTE-HTMLplusTIME-19980918, September 1998.

[WR94] T. Wahl and K. Rothermel. Representing Time in Multimedia Systems. In *Proc. IEEE International Conference on Multimedia Computing and Systems*, pages 538–543, Boston, MA, May 1994.